Touchscreen Biometrics Across Multiple Devices

Tuan Nguyen
New York Institute of Technology
Computer Science Department
tnguye15@nyit.edu

Jonathan Voris
New York Institute of Technology
Computer Science Department
jvoris@nyit.edu

ABSTRACT

As the cost of mobile devices decreases, it is becoming increasingly common for users to own more than one. The presence of multiple pieces of mobile technology complicates the question of how to secure them. Utilizing different authentication solutions on different devices may create usability challenges, while using the same authentication technique on more than one device raises the possibility of a compromise of one device affecting the others. Behavioral biometrics, which model the manner in which users interact with their devices, are an appealing option for a single authentication mechanism solution which is capable of working across different devices. Whether or not a user's behavioral features are specific to a particular device is an open question, however. Intuitively, a user's behavior should be independent of what device they are using. In practice, however, this behavior may be impacted by device hardware and software characteristics such as form factor and virtual keyboard layout.

This paper presents an initial investigation into whether or not biometric touchscreen profiles (i.e., trained classification models which can be utilized to authenticate users to their devices) can be applied across more than one mobile device. We conduct a preliminary IRB-approved investigation in which 10 users were asked to perform 3 common tasks on 3 different mobile devices: reading, typing, and playing a game. We then applied the well-known Support Vector Machine (SVM) learning algorithm to touchscreen features collected during each task. The results of this small-scale study indicate that user behavior is consistent for gameplay and reading across different types of mobile hardware, but different for typing. This provides preliminary evidence that it is possible to apply behavior-based authentication across multiple devices in some, but not all, contexts.

1. INTRODUCTION

Mobile devices are becoming more affordable and are being offered in a variety of form factors, from small smartphones to large tablets and everywhere in between. As a result of these trends, it is no longer uncommon for people to own multiple mobile devices. A recent survey found that 66% of users in the United States own more than one mobile device and 36% own at least three [1]. Another study reported that each consumer will own an average of approximately 2.8 mobile devices by 2020 [9]. Users may have one device for business communication and a different one for personal contacts. Many people also prefer to use a tablet at home and use a smart phones while commuting because of their more portable form factor, for instance.

The increased prevalence of multiple device ownership complicates the process of securing them. Applying different authentication mechanisms on various devices may create usability challenges. The increased cognitive load may make it more difficult for users to recall the specifics of each authentication procedure. On the other hand, using the same authentication process on more than one device raises the possibility of a compromise of one device affecting the others. For example, in order to improve memorability, users may opt to use the same shared authentication secret, such as a password, on more than one device.

Behavioral biometrics, which operate by modeling the manner in which users interact with their devices, are an appealing option for a single authentication mechanism solution which is capable of working across different devices. Previous research has demonstrated behavioral biometrics to be a promising approach to mobile device authentication [3, 18, 11]. With this technique, a biometric profile representing how a user interacts with a particular device modality, such as a touchscreen [8, 6, 19], is established based on features extracted from prior usage. This profile is a trained classification model which can be applied to authenticate a user to his or her device by comparing current device usage against past behavior. Behavioral biometrics offer natural usability benefits since they operate by measuring typical device usage. Using a single biometric profile across all of a user's devices would be beneficial by reducing training times and improving accuracy by providing more data.

Whether or not a user's behavioral features are specific to a particular device is an open question, however. Intuitively, a user's behavior should be independent of what device they are using, and thus their profile should apply across multiple devices. In practice, however, this behavior may be impacted by device hardware and software characteristics such as form factor and virtual keyboard layout. For example, the size of a touchscreen may alter a user's swiping behavior, while different virtual keyboard layouts may affect typing patterns. To demonstrate the potential differences between devices which may impact user behavior, Figure 1 shows the different virtual keyboard used on 3 mobile devices from different manufacturers.

To study this question we performed an Institutional Review Board (IRB) approved preliminary study in which users were asked to perform the same tasks on three different mobile

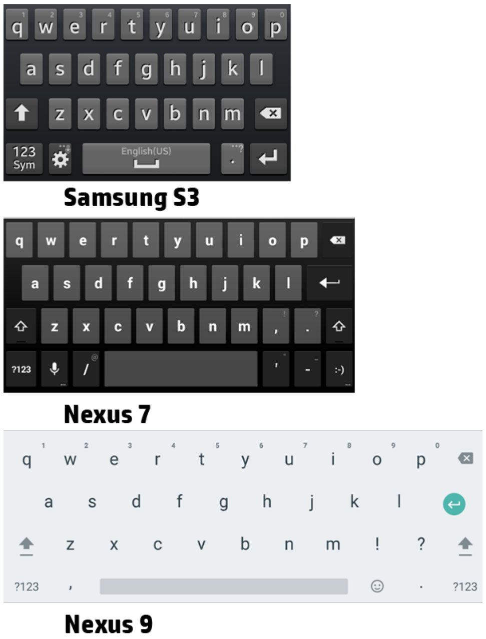

Figure 1: Keyboard Screenshots from the 3 Tested Mobile Devices

devices, each with different form factors. Each of our 10 participants was asked to play a game, read a news article, and write a summary on each of the 3 devices while our sensor application recorded their touchscreen input. We extracted touchscreen behavior features from each users' activity. We applied the popular Support Vector Machine (SVM) learning algorithm to see if we could classify, and therefore authenticate, users irrespective of which device they were using. The results of this initial small-scale study indicate that user behavior is consistent between devices for the reading and game activities, but differences emerged when users performed the writing task on different devices. This provides preliminary evidence which supports that behavioral biometric profiles can be applied across multiple mobile devices in some, but not all, usage scenarios.

2. RELATED WORK

A variety of different authentication mechanisms have been proposed for mobile devices. Most desktop authentication techniques can be applied, but may suffer from usability issues in a mobile context. For instance, it is challenging to enter the long strings of varied characters required for strong passwords on a small mobile device touchscreen [14]. Authentication schemes designed to work well specifically in a mobile setting have been developed to address these issues. Graphical passwords, for instance, are a popular authentication method for mobile devices. They are susceptible to observation attacks [2], however, and may have lower than expected entropy due to predictable user password design choices [5]. Traditional biometrics, which measure a user's physical attributes, are more difficult to lose or steal than knowledge based solutions, but still suffer from some shortcomings. For example, fingerprint reading seems well-suited to mobile authentication, yet is vulnerable to spoofing attacks [13]. Moreover, traditional biometrics often need specialized hardware to function properly, such as a fingerprint reader or iris scanner.

Behavioral biometrics, which operate by analyzing implicit user activities rather than physical characteristics, have recently been studied as a potential solution for a variety of security issues. Perhaps the most well-known application is the use of keystroke [12] and mouse [16] dynamics as a form of authentication for desktop systems. This approach has been adapted to address other security issues as well, such as detecting blog bots to prevent spam and malicious links [4]. Recently, due to the rapid growth of mobile devices in business environments, more research has focused on mobile behavioral authentications to protect sensitive data on mobile devices. Proposed modalities which have shown promise in this area include touchscreen usage [8, 6, 19], graphical touch traces [20], a combination of touch and device movement [3], and application usage habits [18].

A shortcoming of previous work in the area of mobile behavioral biometrics is that studies are performed with all users utilizing a single device throughout the study. It is therefore unclear whether biometric models can be transferred between devices or applied to multiple devices at the same time. We selected our three user tasks for our study - typing [7, 10], reading [8], and playing games [15] - because they had been demonstrated to be used to derive reliable behavioral biometrics by past research.

3. EVALUATION

3.1 Study Design

We conducted a preliminary IRB-approved human subject study with 10 participants recruited from our institution. The experiments were performed with 3 different devices: a Samsung Galaxy S3, Asus Nexus 7, and HTC Nexus 9. We intentionally selected device models from multiple manufacturers for our study because we desired to test whether the subtle changes introduced by differences in device hardware and firmware had an impact on user behavior. For instance, each of the 3 devices had touchscreens with different sizes, resolutions, and sensing capabilities which could potentially have an impact on a user's behavioral touchscreen features. The devices selected for the study have a significant variation in screen size: 4.8 inches, 7 inches, and 8.9 inches for the Galaxy S3, Nexus 7, and Nexus 9 respectively. Each participant was asked to play the mobile game Fruit Ninja [17] as the gameplay task, read a news article in a web browser as a reading task, and write a summary of the article as the writing task. We selected Fruit Ninja as the game play task because it is a highly popular game, having received 100 million downloads, features simple gameplay, and has a short learning curve, and as such is applicable to a broad population.

For the reading task, participants were asked to read articles from a set of recent stories featured in local and national news sources. Each user was presented with the same articles in the same order and asked to read them. For the writing task, we asked users to type summaries of the articles they read in our Android sensor application, which logged their keystrokes, including function keys such as backspace and enter. Each participant was asked to perform each task as they naturally would for a duration 5 minute for a total of 15 minutes of data per user on each device and 45 minutes of collected touchscreen usage data per user overall. Users were provided with a post-conditional questionnaire at the conclusion of the study to collect demographic information and assess how comfortable they felt performing each task on each device.

3.2 Feature Extraction

We developed a sensor application for Android to record all touchscreen interactions between users and the devices while the study was taking place. In the Android security model, applications are executed in separate sandboxes, isolating applications and their data from one another. Our sensor application thus had to rely on direct calls to the Android system command "getevent" in order to acquire raw touchscreen input data. This raw touch data is processed in order to extract high level features which can potentially be used to discriminate between users, which are then used in the behavioral authentication process. Due to differences in device hardware some features are supported in one device but not available in others. For example, the Nexus 7 and Nexus 9 devices record the pressure of a gesture, but the Galaxy S3 did not. For the purposes of this study, only features that were available on all 3 devices were used. When performing the writing task, the standard Android keyboard was used, although the default keyboard differed between devices in terms of key placement and aesthetics such as color, as shown in Figure 1.

We utilized a combination of touchscreen features which were found to be useful for classification in previous research [8] as well as new features we developed. The high level features extracted from each touchscreen gesture are (1) the initial X coordinate of the gesture, (2) the initial Y coordinate of the gesture, (3) the final X coordinate of the gesture, (4) the final Y coordinate of the gesture, (5) the time period during the gesture, (6) the average finger width contacting the screen during the gesture, (7) the length of the gesture along the X axis, (8) the length of the gesture along the Y axis, (9) the distance traveled during the gesture, (10) the direction of the gesture, (11) the speed along X axis of the gesture, (12) the speed along the Y axis of the gesture, (13) the speed along the gesture's trajectory, (14) the velocity of the gesture, (15) the angular velocity of the gesture, (16) the ratio between the length along the X axis and traveled distance of the gesture, (17) the ratio between the length along the Y axis and traveled distance of the gesture, (18) the finger width change during the gesture, (19) the finger width change per time during the gesture, (20) the 8 cardinal and inter-cardinal directions of the gesture, (21) the acceleration of the gesture, (22) the acceleration along X axis of the gesture, and (23) the acceleration along Y axis of the gesture.

3.3 Data Modeling and Analysis

We implemented R language scripts using the "e1071" package to apply a multi-class Support Vector Machine (SVM) to classify users. 300 gestures were randomly selected for each combination of participant, device, and task. To validate the consistency of user behavior on multiple devices, we also combined the data collected on all 3 devices prior to classification. Because each device has a different screen resolution and pixel-per-inch density, we normalized all features affected by those characteristics, including gesture coordinates and finger width. This was done to provide a more accurate basis of comparison between the different devices used in our study. The coordinate range of each device is normalized to a range of 0 to 1 by dividing the original coordinate values by the maximum coordinate on the device. Any features related to user finger characteristics, such as press width and area covered, are converted from pixels to inches. All features used in our experiments are standardized to have a mean of zero and a standard deviation of 1 before modeling to prevent overweighing any particular feature. To achieve multi-class classification, a "one-versus-one" classification technique was employed in which binary classification is applied to each pair of users. 10-fold cross validation was used during model validation.

To measure classification performance we plotted Receiver Operator Characteristic (ROC) curves and calculated the area under ROC curve (AUC). An ROC curve is a plot of a classifier's false positive rate (FPR) on the X axis against its true positive rate (TPR) on the Y axis by adjusting the acceptance threshold used in the classification process. To balance the false accept and false reject rates when assessing classification performance, the Equal Error Rate (EER) of the classifier is considered, which is the common value where the false positive and negative rate are equal.

4. RESULTS

4.1 Survey Results

Responses to our post-conditional questionnaire indicated that our participant pool consisted of university students between the ages of 18 to 34. Half of our subjects were graduate students and the other half were undergraduates; similarly, there was a balance between male (50%) and female (50%) subjects. Though not representative of the broader population due to young age and education level, since our study compared each specific user's behavior between devices, this group's data still functioned as a good basis from which to make a preliminary determination about whether touchscreen biometrics could be modeled across distinct devices. A study with a larger, more diversified participant pool may be pursued as future work. Every participant in the study indicated that they had experience with mobile devices. Moreover, 70% of our subjects owns more than one devices, with each possessing an average of 2.6 devices, confirming the results of surveys of the broader U.S. population [1, 9].

Participants responded that reading and writing activities, which we chose as representative tasks for our study, are two of the most common activities performed on mobile devices. In particular, everyone reported to have used mobile devices to send and receive email, 90% stated they read at least one article on their devices per day, and 70% spent at least 1 hour writing messages on their devices. 90% of participants used mobile devices to play games, with 70% playing games on a daily basis. Our survey also shows the trend of increased deployment of biometrics in authentication. 7 out of 10 subjects use some form of biometrics to unlock their mobile devices, representing the leading authentication method among our participants.

Figure 2 shows responses to our post-conditional survey questions about device responsiveness and the usability of executing our tasks on each device; these questions were posed as five point Likert items. Although the Nexus 9 was found to be the most responsive, users felt that its touchscreen was actually less responsive than the other devices. We attribute

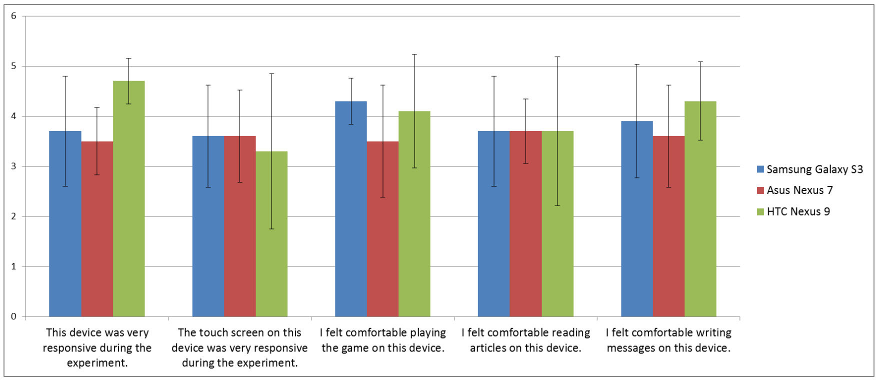

Figure 2: Average Responses to Post-Conditional Questionnaire

this seemingly contradictory response to the perception that the Nexus 9's screen was "slower" due to its larger form factor, which induced delays as users had further to traverse to make inputs equivalent to the other devices. Users indicated that they felt slightly more comfortable using the Nexus 9's larger screen to type than the other devices, but in general, there was not much variation in usability ratings between tasks or devices.

4.2 Classification Results

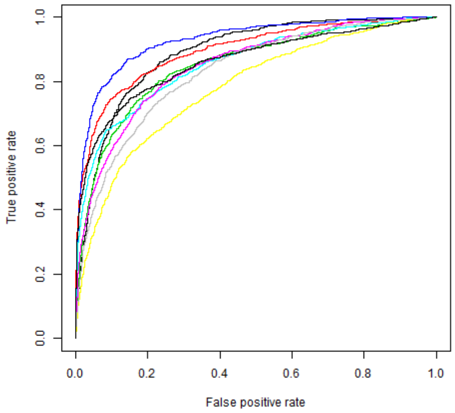

Figure 3: ROC Curves for the Gameplay Task

Our modeling script calculates the classification probability that a validation instance belong to a given user, a process which is applied for each study participant. Classification was first performed for each task on a per-device basis. That is, we applied our classifier to data collected only from the Galaxy S3, then only to usage data from the Nexus 7, then only to data derived from the Nexus 9. After assessing the classification performance on each separate device, we combined each user's behavioral data from all three devices to create a collection of cross-device usage data for each participant. The classification process was then applied to this cross-device usage dataset to determine if the differences between the devices impacted classification performance. This process was repeated for each of the three application tasks (gameplay, reading, and writing) we tested.

We plotted ROC curves for our multi-class SVM classifier by varying the acceptance threshold that is applied to these probabilities. Figure 3, 4, and 5 present ROC curves for the gameplay, reading, and writing activities. We calculated the AUC for each activity based on these curves, which are shown in Table 1. Table 1 also presents the average EER for classifying our study participant's touchscreen features on the test devices. For Fruit Ninja, the average AUC values were 85.68%, 89.77% and 86.48% for the Samsung Galaxy S3, Nexus 7 and Nexus 9, respectively. The average of those

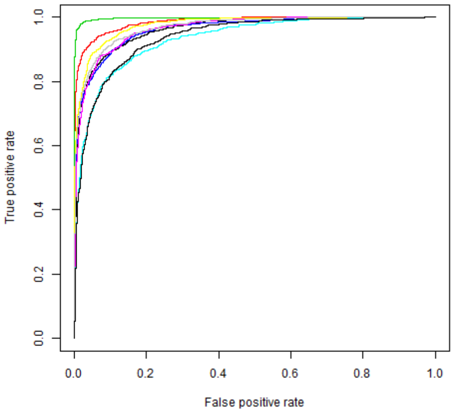

Figure 4: ROC Curves for the Reading Task

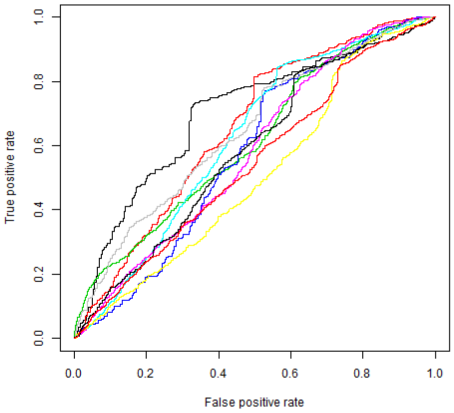

Figure 5: ROC Curves for the Writing Task

Devices	Activity	Average AUC	Average EER
Samsung Galaxy S3	Gameplay	0.8568	22.73%
Asus Nexus 7	Gameplay	0.8977	17.67%
HTC Nexus 9	Gameplay	0.8648	20.79%
Cross-device	Gameplay	0.8573	21.70%
Samsung Galaxy S3	Reading	0.9571	10.87%
Asus Nexus 7	Reading	0.9798	7.09%
HTC Nexus 9	Reading	0.9767	7.24%
Cross-device	Reading	0.9675	8.77%
Samsung Galaxy S3	Writing	0.662	38.20%
Asus Nexus 7	Writing	0.7824	28.51%
HTC Nexus 9	Writing	0.5901	44.44%
Cross-device	Writing	0.6043	43.43%

Table 1: Multi-class SVM Classification Results for All Activities

Activity	F	F-critical
Fruit Ninja	2.1359	2.8663
Reading	0.4759	2.8663
Writing	14.2542	2.8663

Table 2: ANOVA Results for All Activities

value are 87.31%, while the average AUC for classification applied to the data across all 3 devices is a nearly-equal 85.73%. The reading activity showed better classification performance, which is 95.71%, 97.98%, 97.67% and 96.75% for the devices in the same order. Those results are nearly equal as well, with the difference between the worst and the best AUC values being approximately 1%.

These error rates would be unacceptably high for a realistic mobile authentication deployment, but we did not attempt to optimize the performance of our classifier, nor did we experiment with different classification algorithms, reserving these tasks as targets of future work. The purpose of these measurements was to provide a basis for comparison in order to answer the question: does the performance of user classification diminish when behavior data is collected from different devices?

Though the performance of our classifier varied by task - with reading being relatively successful, writing very error prone, and gameplay in between - this information was primarily of interest to our study as a comparison as to whether the classification performance was affected by the use of different devices to a significant degree. To this end, we perform a one-way ANOVA test on these results to determine if there were any statistically significant differences between classification performances on each devices and across all the devices. The null hypothesis is that classification performance is the same for all devices. The results for ANOVA test are shown in Table 2, in which the F and F-critical values of each activity are considered. F is the ratio of the between-group mean square to the within-group mean square. The F-critical values in Table 2 are calculated at a 95% significance level. For the reading and gameplay tasks, because the obtained F value is lower than F- critical ($2.1359 < 2.8663$ for gameplay and $0.4759 < 2.8663$ for reading), we accept the null hypothesis; that is, conclude that there are no differences in the modeling results. This implies that the same biometric model can be used to authenticate users across the different devices we tested as they perform these tasks.

For the writing activity, however, the F value is higher than the F-critical value. Thus, writing does not exhibit the same degree of consistency of user behavior between the different devices. One reason for this is due to the different screen size of each device, which makes the time taken to perform equivalent hand movements from one key to another differ between devices. Differences in keyboard layouts between the devices may contribute to this inconsistency as well. Figure 1 presents screenshots of the default virtual keyboards from the 3 studied devices. We scaled each screenshot to give them the same height for visual comparison purposes to show the difference in the height to width ratio of the keyboards. As shown, the Nexus 9 keyboard is noticeably wider than the rest, for example. The layout of the keys also varies between devices.

We also calculated the effect size of cross-device authentication by calculating Cohen's d for the AUC values obtained from our study. This was accomplished by taking the difference between the mean AUC of classification on each device and the mean AUC of classification across all devices and dividing by the pooled standard deviation for each task. The average effect size was small to medium for the gameplay ($d = 0.374$) and reading tasks ($d = 0.551$) and very large for the writing task ($d = 1.163$), indicating that while the different devices introduced some variations between behavior for each task, these effects were particularly pronounced for the process of typing on each devices' touchscreen.

5. CONCLUSION

As people continue to use additional mobile devices, the need to consider the usability of their security mechanisms becomes more critical. This paper presented an initial investigation into whether touchscreen biometric models can be applied across different mobile devices. We performed a 10 user pilot study in which participants performed 3 tasks on 3 different devices. The results of our study provide preliminary evidence that biometric modeling can successfully be applied across devices for some tasks, such as reading and playing games, but less successfully for others, such as typing. It is therefore important to consider how a device's hardware and software will impact the particular context in which behavioral biometrics will be used. As future work, we intend to pursue a larger scale study of cross-device touchscreen biometric modeling with a broader participant pool, different modeling techniques and features, and additional mobile usage tasks.

6. REFERENCES

[1] M. Anderson. Smartphone, computer or tablet? 36% of Americans own all three. Available at: `http://www.pewresearch.org/fact-tank/2015/11/25/device-ownership/`, 2015.

[2] R. Biddle, S. Chiasson, and P. C. Van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys*, 2012.

[3] C. Bo, L. Zhang, T. Jung, J. Han, X.-Y. Li, and Y. Wang. Continuous user identification via touch and movement behavioral biometrics. In *IEEE International Performance Computing and Communications Conference*, 2014.

[4] Z. Chu, S. Gianvecchio, A. Koehl, H. Wang, and S. Jajodia. Blog or block: Detecting blog bots through behavioral biometrics. 2012.

[5] D. Davis, F. Monrose, and M. K. Reiter. On user choice in graphical password schemes. In *USENIX Security Symposium*, 2004.

[6] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann. Touch me once and i know it's you!: Implicit authentication based on touch screen patterns. In *SIGCHI Conference on Human Factors in Computing Systems*, 2012.

[7] B. Draffin, J. Zhu, and J. Zhang. Keysens: Passive user authentication through micro-behavior modeling of soft keyboard interaction. 2014.

[8] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security*, 2013.

[9] M. Future. The Rise of Mobile: 11.6 Billion Mobile-Connected Devices By 2020. Available at: `http://mobilefuture.org/the-rise-of-mobile-11-6-billion-mobile-connected-devices-by-2020/`, 2016.

[10] H. Gascon, S. Uellenbeck, C. Wolf, and K. Rieck. Continuous authentication on mobile devices by analysis of typing motion behavior. In *Sicherheit*, 2014.

[11] M. Jakobsson, E. Shi, P. Golle, and R. Chow. Implicit authentication for mobile devices. In *USENIX Conference on Hot Topics in Security*, 2009.

[12] K. S. Killourhy and R. A. Maxion. Comparing anomaly-detection algorithms for keystroke dynamics. In *IEEE/IFIP International Conference on Dependable Systems & Networks*, 2009.

[13] S. R. Labs. Fingerprints are not fit for secure device unlocking. Available at: `https://srlabs.de/bites/spoofing-fingerprints/`, 2014.

[14] W. Melicher, D. Kurilova, S. M. Segreti, P. Kalvani, R. Shay, B. Ur, L. Bauer, N. Christin, L. F. Cranor, and M. L. Mazurek. Usability and security of text passwords on mobile devices. In *SIGCHI Conference on Human Factors in Computing Systems*, 2016.

[15] P. Scindia and J. Voris. Exploring games for improved touchscreen authentication on mobile devices. In *Twelfth Symposium on Usable Privacy and Security*, 2016.

[16] C. Shen, Z. Cai, X. Guan, Y. Du, and R. A. Maxion. User authentication through mouse dynamics. *IEEE Transactions on Information Forensics and Security*, 2013.

[17] H. Studios. Fruit Ninja: The Greatest Fruit-Slicing Game in the World. Available at: `https://fruitninja.com`, 2017.

[18] J. Voris, Y. Song, M. B. Salem, and S. Stolfo. You are what you use: An initial study of authenticating mobile users via application usage. In *EAI International Conference on Mobile Computing, Applications and Services*, 2016.

[19] H. Xu, Y. Zhou, and M. R. Lyu. Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones. In *Symposium On Usable Privacy and Security*, 2014.

[20] X. Zhao, T. Feng, and W. Shi. Continuous mobile authentication using a novel graphic touch gesture feature. In *International IEEE Conference on Biometrics: Theory, Applications and Systems*, 2013.